



## **wKinMut: An integrated tool for the analysis and interpretation of mutations in human protein kinases**

**Gonzalez-Izarzugaza, Jose Maria; Vazquez, Miguel; del Pozo, Angela; Valencia, Alfonso**

*Published in:*  
B M C Bioinformatics

*Link to article, DOI:*  
[doi:10.1186/1471-2105-14-345](https://doi.org/10.1186/1471-2105-14-345)

*Publication date:*  
2013

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*  
Gonzalez-Izarzugaza, J. M., Vazquez, M., del Pozo, A., & Valencia, A. (2013). wKinMut: An integrated tool for the analysis and interpretation of mutations in human protein kinases. *B M C Bioinformatics*, 14, [345].  
<https://doi.org/doi:10.1186/1471-2105-14-345>

---

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# wKinMut: An integrated tool for the analysis and interpretation of mutations in human protein kinases

Jose MG Izarzugaza<sup>1,2,\*</sup>

\* Corresponding author

Email: josemgizarzugaza@cbs.dtu.dk

Miguel Vazquez<sup>1</sup>

Email: mvazquezg@cnio.es

Angela del Pozo<sup>1</sup>

Email: adelpozo@cnio.es

Alfonso Valencia<sup>1\*</sup>

\* Corresponding author

Email: avalencia@cnio.es

<sup>1</sup> Structural Biology and BioComputing Programme, Spanish National Cancer Research Centre (CNIO), C/Melchor Fernandez Almagro, 3, E-28029 Madrid, Spain

<sup>2</sup> Department of Systems Biology, Center for Biological Sequence Analysis, Technical University of Denmark, 2800 Lyngby, Denmark

## Abstract

### Background

Protein kinases are involved in relevant physiological functions and a broad number of mutations in this superfamily have been reported in the literature to affect protein function and stability. Unfortunately, the exploration of the consequences on the phenotypes of each individual mutation remains a considerable challenge.

### Results

The wKinMut web-server offers direct prediction of the potential pathogenicity of the mutations from a number of methods, including our recently developed prediction method based on the combination of information from a range of diverse sources, including physicochemical properties and functional annotations from FireDB and Swissprot and kinase-specific characteristics such as the membership to specific kinase groups, the annotation with disease-associated GO terms or the occurrence of the mutation in PFAM domains, and the relevance of the residues in determining kinase subfamily specificity from S3Det. This predictor yields interesting results that compare favourably with other methods in the field when applied to protein kinases.

Together with the predictions, wKinMut offers a number of integrated services for the analysis of mutations. These include: the classification of the kinase, information about associations of the kinase with other proteins extracted from iHop, the mapping of the mutations onto PDB structures, pathogenicity records from a number of databases and the classification of mutations in large-scale cancer studies. Importantly, wKinMut is connected with the SNP2L system that extracts mentions of mutations directly from the literature, and therefore increases the possibilities of finding interesting functional information associated to the studied mutations.

## Conclusions

wKinMut facilitates the exploration of the information available about individual mutations by integrating prediction approaches with the automatic extraction of information from the literature (text mining) and several state-of-the-art databases.

wKinMut has been used during the last year for the analysis of the consequences of mutations in the context of a number of cancer genome projects, including the recent analysis of Chronic Lymphocytic Leukemia cases and is publicly available at [http://wkinmut.bioinfo.cnio.es\\_](http://wkinmut.bioinfo.cnio.es_)

## Background

Current high-throughput resequencing screenings [1-3] represent a powerful set of techniques to discover large numbers of mutations. Of these, only a small fraction are causally implicated in disease onset and therefore, separating the wheat from the chaff is still a major challenge [4]. The interpretation of the overwhelming wealth of data also represents an issue in other fields, such as protein function prediction [5]. For a small subset of the new mutations discovered, experimental information regarding the relationship between the mutation and the underlying biochemical mechanism is known. However, there is no information for the remaining mutations. The intensive requirement of resources makes it unfeasible to experimentally test the association of all these mutations to disease, and to characterize their functional effects. Nevertheless, this problem is very amenable to *in silico* predictors [4,6,7]. Different approaches are currently available to predict the probability of a newly discovered mutation being implicated in disease. Some methods identify crucial positions in a given protein and derive generalized rules to predict the pathogenicity of mutations. Other methods assume that evolutionarily conserved protein residues are important for protein structure, folding and function, whereby mutations in these residues are considered deleterious [8]. Variations on this principle lead to methods that predict deleterious mutations by evaluating changes in evolutionarily conserved PFAM motifs [9]. A number of systems use protein structures to characterize substitutions that significantly destabilize the folded state. There are also methods that integrate prior knowledge in the form of both sequence and structure-related features from a set of experimentally characterized mutations to train automatic machine-learning systems. These systems can infer the pathogenicity of new mutations based on the cases evaluated. Albeit similar in purpose, very different machine-learning methods can be implemented. Among them, probably the most popular ones are: rule-based systems [10-12], decision trees [13], random forests [14,15], neural networks [16,17], Bayesian methods [18] and SVMs [19-23]. Recently, some meta approaches that combine different methodologies have been implemented. For example, Condel [24] integrates five of the most widely employed computational tools for detecting

pathogenic single nucleotide variations. Predictors can also be classified according to their scope. Most of the predictors are generally applicable to amino acid sequences from any protein family, while a few of them include properties that apply only to a given protein family of interest; i.e. protein kinase specific predictors [20,23]. These family-related features bring discriminative information that justifies the development of specialized predictors.

A broad number of mutations in the protein kinase superfamily have been reported in the literature [25] and a subset of them is known to disrupt protein structure and function [26]. For some cases, since human protein kinases are involved in a plethora of physiological functions, this disruption can be causally associated to disease [27]. Still, the majority of protein kinase mutations are tolerated without apparent significant effects [28,29].

In previous publications, we have discussed the preferential distribution of germline pathogenic deviations [30] and driver somatic mutations [31] with respect to regions of functional and structural importance. Here we present, wKinMut, an integrated web-service for the collection of information from multiple sources and for the prediction of the pathogenicity of mutations by combining several prediction approaches. The objective of wKinMut is to provide a one-stop resource for the analysis and interpretation of the consequences of mutations in the protein kinase superfamily.

## Implementation

wKinMut represents the first resource to provide an integrated tool for the analysis and interpretation of the consequences of mutations in the protein kinase superfamily. The main objective of wKinMut is to aid computational biologists and clinicians to prioritize pathogenic mutations and to understand the mechanisms by which some mutations lead to disease, and particularly, to cancer.

The tool presented here, incorporates information retrieval and prediction approaches and displays information from diverse sources. First, it simplifies the collection of information about the mutations, such as the classification, domain architecture, functional annotations and plausible interaction partners of the kinase. Furthermore, kinase mutations are analyzed in their structural context and mentions in dedicated databases, genotyping studies and the literature that suggest an implication in disease are also presented. Second, wKinMut estimates the theoretical pathogenicity of kinase mutations with three different approaches, including our newly developed kinase-specific method, KinMut [23], based on the evaluation of a wide set of sequence-derived features that describe each independent mutation. The affected domain and kinase group, diverse functional annotations, residue physicochemical properties and relevance of the mutated residues in determining subfamily specificity are considered.

wKinMut has been implemented mostly in Ruby. The functionality is implemented as a workflow accessible through a REST interface that can render the results either in JSON format or HTML. The later constitutes the interface described in this document. Some of data resources that support this system, such as gene descriptions or iHOP interactions, are queried remotely through the internet as demanded; but are then cached to improve subsequent accesses. The server incorporates some additional caching schemes to improve performance in the back-end, by persisting the job results, and in the web interface, by caching the HTML.

## Web interface

### *Step 1: submission of mutations for analysis*

The input to wKinMut are non-synonymous mutations in the protein kinase superfamily. The input format should encode the Uniprot/Swissprot accession number, the wild type residue, the position and the mutated residue. Non-standard amino acids and truncating mutations will be excluded from the analysis. An example of this format would be a mutation from Glycine to Alanine in position 719 of the human epidermal growth factor receptor, which is encoded as P00533 G719A. In the following sections, we will use this example to guide the reader through the different result views (Figure 1). Multiple mutations can be submitted at a time, either as a plain text file or directly via the applications form, the sample dataset provided as part of wKinMut's documentation can be used as a formatting guide.

---

**Figure 1 Summary of the different result pages in wKinMut: Example of a Gly-719-Ala mutation in the human epidermal growth factor receptor.** The figure shows an example of an input to the server (panel **a**) and the results summary table (panel **b**). The rest of panels display show the different outputs from the server, including the gene/protein summary tab (panel **c**), the domain tab (panel **d**), the structure view (panel **e**), the pathogenicity assessment (panel **f**). Information from the databases, the literature and iHop is exemplified in panels **g**, **h** and **i** respectively.

---

### *Step 2: interpretation of the consequences of the mutations*

The first output the user will get right after submitting the mutations is a summary page with useful information about the requested mutations (Figure 1, panel b). It includes a description of the proteins in Uniprot, the membership to kinase groups in the classification in KinBase [32,33] and the estimation of the pathogenicity of mutations attending to our kinase-specific predictor of pathogenicity, KinMut [23]. The prediction of the pathogenicity will be discussed in detail in a forthcoming section, nevertheless we decided to include this information at this step as a guide to prioritize mutations. It might be interesting to point out here that users interested only in the results from KinMut, can find a link to the predictions in this summary page that can be accessed programmatically. The scope of wKinMut goes beyond providing raw prediction of pathogenicity from KinMut, the web-service's main goal is to aid computational biologists and clinicians to understand and to interpret the consequences of kinase mutations. Hence, information complementary to KinMut predictions, is provided. In the summary table, the 'View' link in the right-most 'Details' column (Figure 1, panel b) will redirect the user to another page containing this complementary information, which includes: the values of the features used for classification, PFAM domains affected by the mutation, protein-protein interaction information extracted from the literature with iHop [34], mentions of the mutations in the literature automatically mined with SNP2L [25,35], and existing records of the mutations in other dedicated databases. This additional information is intended to provide the basic background to help to understand and interpret the consequences of the mutations. Each individual piece of information will be discussed thoroughly in the following sections.

## *General information about the protein/gene*

Information under the ‘gene/protein’ tab (Figure 1, panel c) focuses on information shared by all mutations in the same kinase. Background information such as the gene name, the formal description in Uniprot and the classification in KinBase [32,33] of the kinase is provided. In addition, the system provides the Gene Ontology terms with which the kinase has been annotated in each of the independent sub-ontologies (namely Molecular Function, Cellular Compartment and Biological Process). This information provides clues to unveil the function of the kinase and it is used by KinMut to calculate the likeness of the protein (and subsequently the mutation) to play a role in disease.

## *PFAM domains*

In a previous publication [23] we demonstrated that mutations occurring in certain domains such as the Tyrosine kinase domain (PKinase Tyr, according to PFAM) are more likely to cause disease. This is coherent with the assumption that the function of some domains is more important than the function of others. In wKinMut, this information is contained in the ‘PFAM domains’ tab (Figure 1, panel d), which displays the domain (or domains, in some cases) where the mutation is occurring and the alignment used by PFAM as seed to generate the domain family. The alignment is evaluated in terms of sequence conservation. Under the assumption that conserved regions have been preserved by evolution, this information can help the user to identify important regions in the structure of the domain.

## *Mapping the mutations onto structures*

To understand the consequences of mutations might have in protein stability and function it is sometimes useful to study the mutations in their structural contexts. However, mapping mutations from sequences to structures is not always trivial [36]. Under the ‘Structures’ tab, wKinMut enables the visualization of the mutation mapped to all available structures. (Figure 1, panel e). In addition, the versatility of the Jmol applet implemented in wKinMut allows advanced users to adapt the visualization to their specific needs.

## *Prediction of the pathogenicity*

In wKinMut the theoretical pathogenicity of mutations is assessed by two independent methods, namely SIFT [8] and KinMut [23]. This information is displayed in the ‘Pathogenicity’ tab (Figure 1, panel f). SIFT [8] predicts whether non-synonymous mutations are prone to affect protein function. This prediction is based on the degree of conservation of the residues in sequence alignments derived from closely related sequences. A threshold value of 0.05 is used to determine that mutations are likely to be pathogenic. KinMut [23] is a kinase-specific predictor of the pathogenicity of mutations. It relies in a machine-learning approach (SVM) to evaluate a number of sequence-derived features that describe kinase mutations from different perspectives, including: a) at the gene level, the membership to a Kinbase group and Gene Ontology terms. b) at the domain level, the occurrence of the mutation inside a PFAM domain, and c) at the residue level, several properties including amino acid type, functional annotations from Swissprot and FireDB [37], specificity-determining positions, etc. SVM scores greater than -0.5 indicate that the mutation is very likely pathogenic. The values of these features are also displayed in this section of the web-

service to aid to interpret the predictions. Please, refer to the original publications for information on the individual characteristics, capabilities and validation of each predictor.

### *Mutations in databases*

The wealth of knowledge provided by current research is usually stored in databases. A number of them store information about mutations from diverse perspectives. In wKinMut (Figure 1, panel g) we collect information from four different sources (namely the Uniprot Variant Pages [38], KinMutBase [39], SAAPdb [26] and COSMIC [40]) in an attempt to cover all aspects of protein kinase mutation. The information displayed includes information about the structural consequences of mutations, experiments associating mutations with a certain disease, or the proof that a mutation has been observed in a cancer sample.

### *Automatic extraction of mutations from the literature*

Unfortunately, the databases referred in the previous section do not contain all current knowledge about mutations. Even in the cases where a database record exists, the knowledgebase cannot always store all contextual information. The context is sometimes very important for the correct interpretation of the predictions: experimental conditions, patients' habits and clinical histories, etcetera. wKinMut provides pointers to mentions of the mutations in the literature under the 'Literature' tab (Figure 1, panel h). We extract this information automatically using our in-house text mining approach, SNP2L [25]. In brief, SNP2L is a literature mining pipeline for the automatic extraction and disambiguation of singlepoint mutation mentions from both abstracts as well as full text articles, followed by a sequence validation check to link mutations to their corresponding kinase protein sequences.

### *Automatic determination of interaction partners*

wKinMut integrates Protein-Protein Interactions (PPI) gathered from iHOP in the homonymous tab (Figure 1, panel i). Briefly, iHOP is a powerful text mining system to automatically extract protein protein interactions from PubMed abstracts. To relate the interaction information with its context, the sentences including the interaction mentions are also provided.

## **Conclusion**

wKinMut facilitates the exploration of the information available about individual mutations by integrating prediction approaches with the automatic extraction of information from the literature (text mining) and several currently available databases. wKinMut works as an open accessible web server.

The system offers direct prediction of the potential pathogenicity of the mutations from a number of methods, including our recently developed prediction method based on the combination of information from a range of diverse sources with a machine learning system [23]. The features used by our new prediction system include: general physicochemical properties, annotations of known functional sites from FireDB and Swissprot and kinase-specific characteristics such as membership to a specific group of kinases, annotations of disease associations extracted from GO terms and mapping of PFAM domains, and relevance of the residues for the differences between groups of kinases. In addition to the predictions,

wKinMut offers a number of integrated complementary services that help to understand the consequences and the mechanism of the mutations. These services include the classification of the kinase, information about associations of the kinase with other proteins extracted directly from the Medline abstracts, the mutations on the corresponding protein structures, and possible relations with pathogenicity recorded in disease-variation databases and from large-scale cancer studies. An important component of wKinMut is the access to information about the mutations extracted directly from the literature. This information is important for the contextualization of the consequences of the mutations. wKinMut uses our previously developed SNP2L [25], that has been shown to provide a substantial addition to the information provided by public databases and repositories..

In summary, we think that wKinMut constitutes a powerful one-stop shop for the study of the potential pathogenic potential of mutations in protein kinases. As such, wKinMut will be of interest for bioinformaticians and computational biologists that can use the information provided by the server programmatically as part of their own analysis pipelines, and it can be also useful to biologists and clinicians who can browse and explore punctual information easily from the provided interface. We have used wKinMut during the past year for the analysis of the consequences of mutations in the context of a number of personalized cancer genome projects (see [41]), including the recent analysis of Chronic Lymphocytic Leukemia cases [42,43].

A further development of the presented system would consider the analysis of the downstream consequences of mutations in relation to potential and known post-translational modifications and their interrelations (see [44,45]). We are interested in extending wKinMut capabilities to the analysis of the combined effect of mutations in pathways and signalling networks in where kinases are essential components wKinMut is publicly available at <http://wkinmut.bioinfo.cnio.es>

## **Competing interests**

The authors declare that they have no competing interests.

## **Author's contributions**

AV and JMGI designed the experiment. MV, JMGI designed the web server. MV, JMGI and AP implemented the web server. JMGI, MV and AV wrote the paper. All the authors read and approved the manuscript.

## **Acknowledgements**

The authors thank the members of the Structural Biology and Biocomputing Programme (CNIO), especially A. Rausell, D. Juan, I. Ezkurdia and T. Pons, for interesting discussion and comments on this manuscript. This research was supported by OpenPhacts European project (115191-2) and Spanish Ministry of Science and Innovation project BIO2007-6685



## References

1. Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JKV, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE: **The consensus coding sequences of human breast and colorectal cancers.** *Science* 2006, **314**(5797):268–274.
2. Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, Shen D, Boca SM, Barber TD, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyansky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JKV, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PVK: **The genomic landscapes of human breast and colorectal cancers.** *Science* 2007, **318**(5853):1108–13.
3. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jenkinson A, Jones D: **Patterns of somatic mutation in human cancer genomes.** *Nature* 2007, **446**(7132):153–8.
4. Baudot A, Real F, Izarzugaza J, Valencia A: **From cancer genomes to cancer models: bridging the gaps.** *EMBO Rep* 2009, **10**(4):359–66.
5. Friedberg I, Jambon M, Godzik A: **New avenues in protein function prediction.** *Protein Sci* 2006, **15**(6):1527–1529.
6. Karchin R: **Next generation tools for the annotation of human SNPs.** *Brief Bioinformatics* 2009, **10**:35–52.
7. Cline M, Karchin R: **Using bioinformatics to predict the functional impact of SNVs.** *Bioinformatics* 2010, **27**(4):441–8.
8. Ng PC, Henikoff S: **Predicting deleterious amino acid substitutions.** *Genome Res* 2001, **11**(5):863–874.
9. Clifford RJ, Edmonson MN, Nguyen C, Buetow KH: **Large scale analysis of non-synonymous coding region single nucleotide polymorphisms.** *Bioinformatics* 2004, **20**(7):1006–1014.
10. Reva B, Antipin Y, Sander C: **Predicting the functional impact of protein mutations: application to cancer genomics.** *Nucleic Acids Res* 2011, **39**(17):e118.
11. Ramensky V, Bork P, Sunyaev S: **Human non-synonymous SNPs: server and survey.** *Nucleic Acids Res* 2002, **30**(17):3894–3900.
12. Wang Z, Moult J: **SNPs, protein structure, and disease.** *Hum Mutat* 2001, **17**(4):263–270.

13. Krishnan VG, Westhead DR: **A comparative study of machine-learning methods to predict the effects of single nucleotide polymorphisms on protein function.** *Bioinformatics* 2003, **19**(17):2199–2209.
14. Kaminker JS, Zhang Y, Waugh A, Haverty PM, Peters B, Sebisanoovic D, Stinson J, Forrest WF, Bazan JF, Seshagiri S, Zhang Z: **Distinguishing cancer-associated missense mutations from common polymorphisms.** *Cancer Res* 2007, **67**(2):465–473.
15. Wainreb G, Ashkenazy H, Bromberg Y, Starovolsky-Shitrit A, Haliloglu T, Ruppin E, Avraham KB, Rost B, Ben-Tal N: **MuD: an interactive web server for the prediction of non-neutral substitutions using protein structural data.** *Nucleic Acids Res* 2010, **38**(Suppl):W523–W528.
16. Ferrer-Costa C, Orozco M, de la Cruz X: **Characterization of disease associated single amino acid polymorphisms in terms of sequence and structure properties.** *J Mol Biol* 2002, **315**(4):771–786.
17. Bromberg Y, Rost B: **SNAP: predict effect of non-synonymous polymorphisms on function.** *Nucleic Acids Res* 2007, **35**(11):3823–3835.
18. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR: **A method and server for predicting damaging missense mutations.** *Nat Methods* 2010, **7**(4):248–249.
19. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R: **Functional annotations improve the predictive score of human disease-related mutations in proteins.** *Hum Mutat* 2009, **30**(8):1237–1244.
20. Torkamani A, Schork NJ: **Accurate prediction of deleterious protein kinase polymorphisms.** *Bioinformatics* 2007, **23**(21):2918–2925.
21. Yue P, Li Z, Moult J: **Loss of protein structure stability as a major causative factor in monogenic disease.** *J Mol Biol* 2005, **353**(2):459–473.
22. Karchin R, Diekhans M, Kelly L, Thomas DJ, Pieper U, Eswar N, Haussler D, Sali A: **LS-SNP: large-scale annotation of coding non-synonymous SNPs based on multiple information sources.** *Bioinformatics* 2005, **21**(12):2814–2820.
23. Izarzugaza JM, Pozo A, Vazquez M, Valencia A: **Prioritization of pathogenic mutations in the protein kinase superfamily.** *BMC Genomics* 2012, **13**(Suppl 4):S3.
24. Gonzalez-Perez A, Lopez-Bigas N: **Improving the assessment of the outcome of non synonymous SNVs with a consensus deleteriousness score.** *Condel. Am J Hum Genet* 2011, **88**(4):440–449.
25. Krallinger M, Izarzugaza JMG, Rodriguez-Penagos C, Valencia A: **Extraction of human kinase mutations from literature, databases and genotyping studies.** *BMC Bioinformatics* 2009, **10**(Suppl 8):S1.

26. Hurst J, McMillan L, Porter C, Allen J, Fakorede A, Martin A: **The SAAPdb web resource: A large-scale structural analysis of mutant proteins.** *Hum Mutat* 2009, **30**(4):616–24.
27. Lahiry P, Torkamani A, Schork NJ, Hegele RA: **Kinase mutations in human disease: interpreting genotype-phenotype relationships.** *Nat Rev Genet* 2010, **11**:60–74.
28. Greenman C, Wooster R, Futreal PA, Stratton MR, Easton DF: **Statistical analysis of pathogenicity of somatic mutations in cancer.** *Genetics* 2006, **173**(4):2187–2198.
29. Stratton MR, Campbell PJ, Futreal PA: **The cancer genome.** *Nature* 2009, **458**(7239):719–724.
30. Izarzugaza JMG, McMillan LEM, Baresic A, Orengo CA, Martin ACR, Valencia A: **Characterization of pathogenic germline mutations in human Protein Kinases.** *BMC Bioinformatics* 2011, **12**(Suppl 4):S1.
31. Izarzugaza J, Redfern O, Orengo C, Valencia A: **Cancer-associated mutations are preferentially distributed in protein kinase functional sites.** *Proteins* 2009, **77**(4):892–903.
32. Manning G, White DB, Martinez R, Hunter T, Sudarsanam S: **The protein kinase complement of the human genome.** *Science* 2002, **298**(5600):1912–1934.
33. Miranda-Saavedra D, Barton G: **Classification and functional annotation of eukaryotic protein kinases.** *Proteins* 2007, **68**(4):893–914.
34. Hoffmann R, Valencia A: **Implementing the iHOP concept for navigation of biomedical literature.** *Bioinformatics* 2005, **21**(Suppl 2):ii252–ii258.
35. Krallinger M, Valencia A, Hirschman L: **Linking genes to literature: text mining, information extraction, and retrieval applications for biology.** *Genome Biol* 2008, **9**(Suppl 2):S8.
36. Izarzugaza JMG, Baresic A, McMillan LEM, Yeats C, Clegg AB, Orengo CA, Martin ACR, Valencia A: **An integrated approach to the interpretation of single amino acid polymorphisms within the framework of CATH and Gene3D.** *BMC Bioinformatics* 2009, **10**(Suppl 8):S5.
37. Lopez G, Valencia A, Tress ML: **FireDB—a database of functionally important residues from proteins of known structure.** *Nucleic Acids Res* 2007, **35**:D219–D223. Database issue.
38. Yip YL, Lachenal N, Pillet V, Veuthey AL: **Retrieving mutationspecific information for human proteins in UniProt/Swiss-Prot Knowledgebase.** *J Bioinform Comput Biol* 2007, **5**(6):1215–1231.
39. Ortutay C, Valiaho J, Stenberg K, Vihinen M: **KinMutBase: a registry of disease-causing mutations in protein kinase domains.** *Hum Mutat* 2005, **25**(5):435–442.

40. Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, Flanagan A, Teague J, Futreal PA, Stratton MR, Wooster R: **The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website.** *Br J Cancer* 2004, **91**(2):355–358.
41. Valencia A, Hidalgo M: **Getting personalized cancer genome analysis into the clinic: the challenges in bioinformatics.** *Genome Medicine* 2012, **4**:61.
42. Quesada V, Conde L, Villamor N, Ordóñez GR, Jares P, Bassaganyas L, Ramsay AJ, Beà S, Pinyol M, Martínez-Trillos A, López-Guerra M, Colomer D, Navarro A, Baumann T, Aymerich M, Rozman M, Delgado J, Giné E, Hernández JM, González-Díaz M, Puente DA, Velasco G, Freije JM, Tubío JM, Royo R, Gelpí JL, Orozco M, Pisano DG, Zamora J, Vázquez M, *et al*: **Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia.** *Nat Genet* 2011, **44**(1):47–52.
43. Puente XS, Pinyol M, Quesada V, Conde L, Ordóñez GR, Villamor N, Escaramis G, Jares P, Beà S, González-Díaz M, Bassaganyas L, Baumann T, Juan M, López-Guerra M, Colomer D, Tubío JM, López C, Navarro A, Tornador C, Aymerich M, Rozman M, Hernández JM, Puente DA, Freije JM, Velasco G, Gutiérrez-Fernández A, Costa D, Carrió A, Guijarro S, Enjuanes A, *et al*: **Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia.** *Nature* 2011, **475**(7354):101–5.
44. Minguéz P, Parca L, Diella F, Mende DR, Kumar R, Helmer-Citterich M, Gavin AC, van Noort V, Bork P: **Deciphering a global network of functionally associated post-translational modifications.** *Mol Syst Biol* 2012, **8**:599.
45. Beltrao P, Albanese V, Kenner L, Swaney DL, Burlingame A, Villen J, Lim WA, Fraser JS, Frydman J, Krogan NJ: **Systematic Functional Prioritization of Protein Posttranslational Modifications.** *Cell* 2012, **150**:413–425.